



**QUEEN'S
UNIVERSITY
BELFAST**

Small RNAs from plants, bacteria and fungi within the order Hypocreales are ubiquitous in human plasma

Beatty, M., Guduric-Fuchs, J., Brown, E., Bridgett, S., Chakravarthy, U., Hogg, R. E., & Simpson, D. A. (2014). Small RNAs from plants, bacteria and fungi within the order Hypocreales are ubiquitous in human plasma. *BMC Genomics*, 15, [933]. <https://doi.org/10.1186/1471-2164-15-933>

Published in:
BMC Genomics

Document Version:
Publisher's PDF, also known as Version of record

Queen's University Belfast - Research Portal:
[Link to publication record in Queen's University Belfast Research Portal](#)

Publisher rights

© 2014 Beatty et al.; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.

General rights

Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact openaccess@qub.ac.uk.

RESEARCH ARTICLE

Open Access

Small RNAs from plants, bacteria and fungi within the order *Hypocreales* are ubiquitous in human plasma

Meabh Beatty, Jasenka Guduric-Fuchs, Eoin Brown, Stephen Bridgett, Usha Chakravarthy, Ruth Esther Hogg and David Arthur Simpson*

Abstract

Background: The human microbiome plays a significant role in maintaining normal physiology. Changes in its composition have been associated with bowel disease, metabolic disorders and atherosclerosis. Sequences of microbial origin have been observed within small RNA sequencing data obtained from blood samples. The aim of this study was to characterise the microbiome from which these sequences are derived.

Results: Abundant non-human small RNA sequences were identified in plasma and plasma exosomal samples. Assembly of these short sequences into longer contigs was the pivotal novel step in ascertaining their origin by BLAST searches. Most reads mapped to rRNA sequences. The taxonomic profiles of the microbes detected were very consistent between individuals but distinct from microbiomes reported at other sites. The majority of bacterial reads were from the phylum *Proteobacteria*, whilst for 5 of 6 individuals over 90% of the more abundant fungal reads were from the phylum *Ascomycota*; of these over 90% were from the order *Hypocreales*. Many contigs were from plants, presumably of dietary origin. In addition, extremely abundant small RNAs derived from human Y RNAs were detected.

Conclusions: A characteristic profile of a subset of the human microbiome can be obtained by sequencing small RNAs present in the blood. The source and functions of these molecules remain to be determined, but the specific profiles are likely to reflect health status. The potential to provide biomarkers of diet and for the diagnosis and prognosis of human disease is immense.

Keywords: Small RNAs, Fungi, Plasma, Microbiome, Metagenomics, Next generation sequencing, MicroRNA, Biomarker, Blood, Y RNA

Background

It has been estimated that there are at least ten times more microbial cells associated with our bodies than there are human cells [1,2]. Recent advances in high throughput, metagenomic sequencing approaches have facilitated identification of this diverse population of microbes at the genomic level. Characterisation of this microbiome, led by the Human Microbiome Project [3], has revealed that its composition varies widely between body sites and between individuals [2,4-7].

The microbiome has a significant influence upon health. The majority of microbes are found in the gut and have essential roles in normal human physiology and immune responses [1,8]. The composition of the gut microbiome is correlated with diet [9] and may be linked with the pathophysiology of bowel disorders [10,11], obesity [12-14], atherosclerosis [15-17], diabetes [18], rheumatoid arthritis [19,20] and neurodevelopmental disorders [21]. Inflammatory bowel conditions have been linked with the intestinal fungal community [4,22].

Most metagenomic studies to date have involved isolation of DNA from external body sites or from the respiratory or digestive tracts, with fecal samples being the most commonly used source for investigation of the gut

* Correspondence: David.Simpson@qub.ac.uk
Centre for Experimental Medicine, Queen's University Belfast, Belfast, Northern Ireland, UK

microbiome. Certain small RNAs are stable in the blood and in particular microRNAs have been widely studied as potential predictors of disease [23,24]. However, we and others [25-27] have observed the existence of additional, exogenous small RNAs of potential microbial origin. Indeed, Wang *et al.* have documented the existence of RNA from bacteria and fungi in plasma and suggested that they may serve as signaling molecules or indicators of human health [25]. The origin of these small RNAs is unclear, but they are almost certainly derived from microbes inhabiting the gut or respiratory tract, rather than from viable microbes within the circulation. Nonetheless, it seems likely that the subset of the total human microbiome which contributes to these blood-borne small RNAs is linked with health status. The ability to reliably determine the composition of this microbiome from the sequences of the small RNAs present in a blood sample could form the basis of an extremely valuable diagnostic test.

The aim of this study was to construct a profile of the microbiome from which the exogenous small RNAs present in human plasma are derived. The merging of overlapping sequences to generate contigs facilitated identification of the origin of the short RNA sequences. The microbiome profiles generated were consistent across 6 individuals (3 from this study and 3 from publicly available data [28]). In addition to bacterial sequences, a large proportion of reads matched fungal sequences. To our surprise, the majority of these were assigned to the order *Hypocreales*. This work has further demonstrated the feasibility of generating a microbiome profile from small RNAs in plasma [25]. The ease of obtaining blood samples will facilitate analysis of this microbiome in a wide range of physiological and disease conditions. These findings also raise the intriguing questions of whether these exogenous RNAs have any functional implications and why sequences from one fungal order are so abundant.

Results

RNA was extracted from three plasma samples and small RNA libraries prepared using an Illumina kit. Each library was sequenced on a MiSeq (Illumina). The unique reads and raw sequencing data have been deposited in Gene Expression Omnibus (GEO), accession number GSE52981. Sequencing data for three plasma exosomal small RNA libraries prepared with a kit from Bioo Scientific were downloaded from GEO [28]. For one of these samples data from libraries prepared with an NEB kit and an Illumina kit (as used in this study) were also available. The strategy for analysis of the sequencing data was to filter out reads derived from human genes, assemble the remaining reads into contigs, annotate these by alignment to known sequences and perform a phylogenetic classification (Figure 1).

The proportions of reads annotated to human genes are illustrated in Figure 2A (absolute numbers in Additional file 1). As expected, a large proportion of reads represented microRNAs, but remarkably, in the whole plasma samples prepared in this study, a similar proportion mapped to Y RNAs. Y RNAs are small cytoplasmic non-coding RNAs that can be cleaved to form smaller RNAs independently of the microRNA pathway [29]. The vast majority of reads (>99%) mapped to hY4, with small numbers to hY5, hY3 and hY1. A smaller but significant number of Y RNA sequences were present in the plasma exosome samples. In small RNA sequencing datasets from whole blood, which included cellular RNAs (GEO accession GSE46579), hY4-derived RNAs were present at levels comparable to an abundant microRNA [30]. The differences in Y RNA abundance observed between studies can be attributed to differences in sample collection (eg whole plasma or plasma exosomes) and library preparation, which result in differing distributions of small RNA read lengths (Additional file 2: Figure S1). The small RNAs detected corresponded to the 5p and 3p arms of the predicted secondary structure of hY4 (Figure 3A). Taqman small RNA RT-qPCR assays employ a stem-loop reverse transcription primer and are therefore expected to be specific for the target small RNA and not detect the full length precursor RNA. Therefore the low Cp values observed with the assays targeting the most abundant hY4 sequences from each arm both confirmed the presence of these small RNAs in plasma and suggested that they are indeed much more abundant than any individual microRNA (Figure 3B). To further confirm the presence of hY4 fragments, RNA was polyadenylated, reverse transcribed with an oligo-dT adaptor and PCR performed with primers specific for the putative hY4 fragments. The size of the product amplified using the 5p primer was consistent with presence of the small RNA template detected in the sequencing rather than full length hY4 RNA (Figure 3C).

A significant number of unannotated reads remained in all samples. The randomly cloned DNA sequences obtained in conventional metagenomic studies are typically assembled into contigs to enhance identification of homology with known genes. Although this strategy would not be applicable to discretely processed small RNAs, such as microRNAs, we reasoned that it could aid detection of longer RNAs which are processed to generate multiple small overlapping RNAs. All the unannotated reads were therefore pooled and assembled into 41542 contigs. For annotation purposes, the 5142 contigs with significant hits ($E < 1 \times 10^{-3}$) in a megablast search of the NCBI non-redundant database were assigned the identity of the top hit (lineages listed in Additional file 3). The unannotated reads from each sample were realigned to these contigs and the proportions of reads

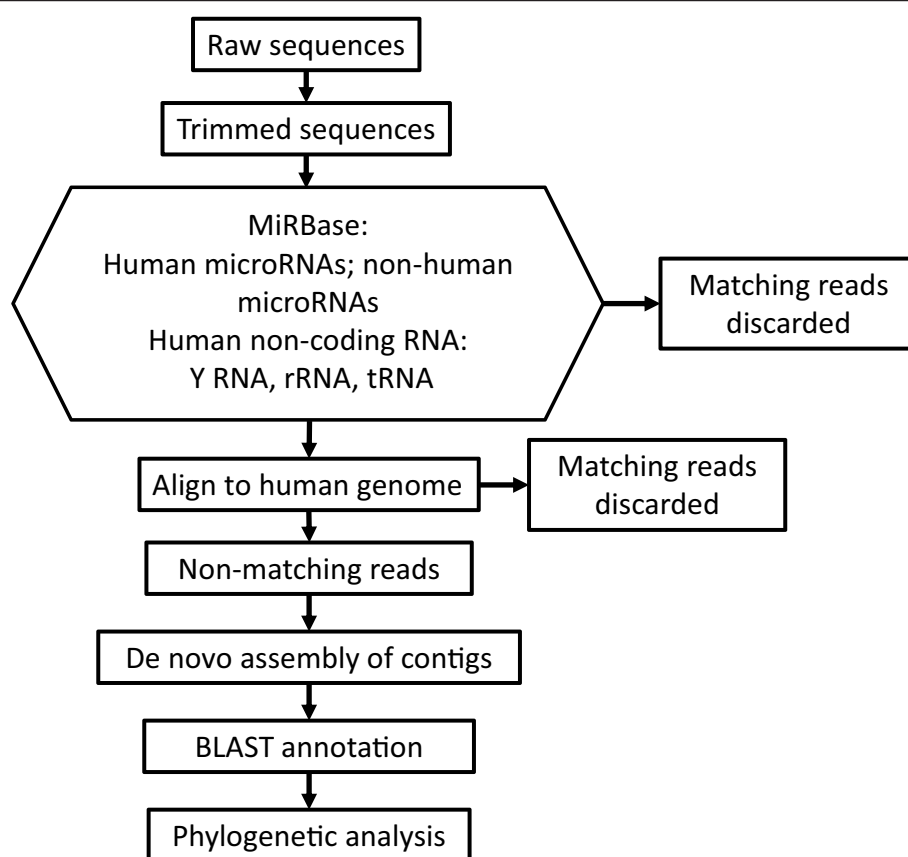


Figure 1 Schema of the strategy for analysis of sequencing data. Reads that did not align to human sequences or other known microRNAs were assembled into contigs. These were annotated by BLAST alignment to the NCBI nr database and phylogenetic analysis performed with the gi numbers of the top resulting hits.

mapping to different taxonomic categories calculated (Figure 2B-F). Most identifiable reads were assigned to Metazoa, Bacteria or Fungi. Although some metazoan reads could be derived from food [31], many are likely to be misassigned due to similarity with human sequences.

A small percentage of contigs matched plant sequences, but due to the conservation of rRNA across the kingdom *Viridiplantae*, the top blast hits did not reliably identify their source, but rather reflected the composition of the database (a preponderance of algal sequences was observed). However, in most instances the sequences were sufficiently divergent from human rRNA to support the notion that they are derived from dietary plant material (Figure 4).

The phylogenetic profile of the bacterial microbiome was remarkably similar between individuals (Figure 2C), with *Proteobacteria* being the most abundant phylum. This is consistent with an origin in the gut. The number of reads matching fungal sequences was higher than expected and of these, more than 90% in 5 of 6 individuals were from the phylum *Ascomycetes* (Figure 2D). Remarkably, it was possible to further define the origin of almost

all these reads to within the class *Sordariomycetes* and order *Hypocreales* (Figure 2E-F). The predominance of sequences from the *Hypocreales* is illustrated when the numbers of reads mapping to each fungal order are placed on a phylogenetic tree comprising all orders with at least one matching contig (Figure 5).

For the 20 exogenous contigs represented by the most reads, the top 5% of BLAST hits (min score 50) were analysed with the MEGAN taxonomic classification tool [32,33]. They all mapped to rRNA, 16 of the 20 to fungal sequences, with the lowest common taxonomic rank for 5 of the top 6 being the fungal order *Hypocreales* or lower (Figure 6). The relative abundances of contigs across the samples were very consistent. Contig 44, which mapped to *Hypocreales* rRNA, was the most abundant in 5 of the 6 individuals. Notably 9 of the top BLAST hits for the 20 contigs were to the genus *Fusarium*. The mycoprotein Quorn is derived from *Fusarium venenatum* [34]. Although it is intriguing to speculate that the sequences we observe are derived from Quorn, it seems unlikely that all 6 subjects would have had this in their diet. In addition, although several contigs align very closely with published

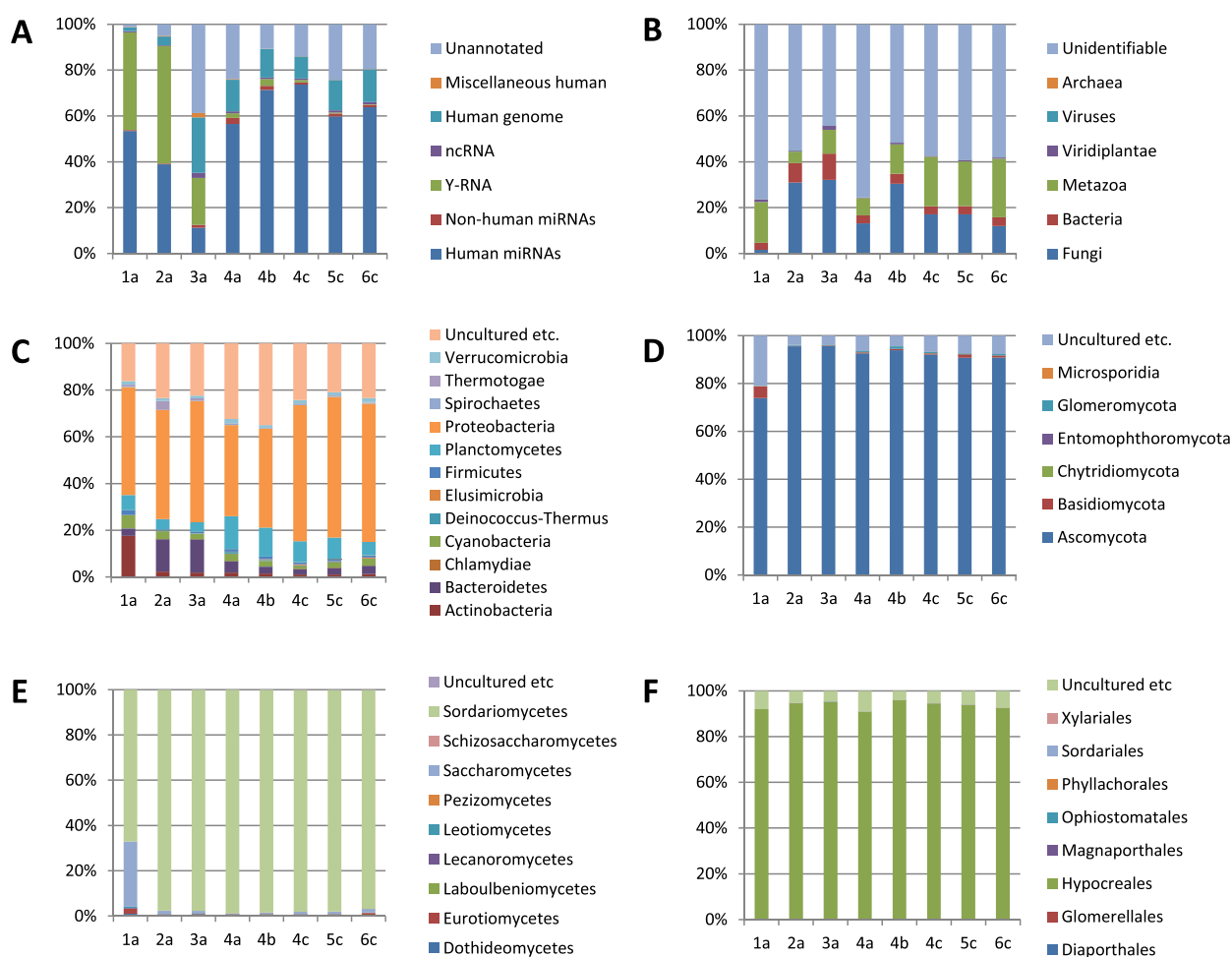


Figure 2 Distribution of human reads by gene type and other reads by organism. Each individual is represented by a number: 1–3 this study (whole plasma); 4–6 Huang et al.[28] (plasma exosomal RNAs). The library preparation method is indicated as follows: a = Illumina; b: NEB; c = Bioo Scientific). (A) 100% stacked columns illustrating the proportions of reads annotated to human genes, non-human microRNAs or unannotated. (B) The proportions of unannotated reads (from (a)) subsequently assigned to superkingdom or kingdom. (C) Bacterial reads assigned to Phyla (those comprising <0.5% in all samples are not illustrated). (D) The proportions of fungal reads by phyla. (E) The proportions of reads assigned to classes within the phylum Ascomycetes. (F) The proportions of reads assigned to orders within the class Sordariomycetes.

F. venenatum rRNA sequences, they match even more closely to other species (Additional file 4: Figure S2).

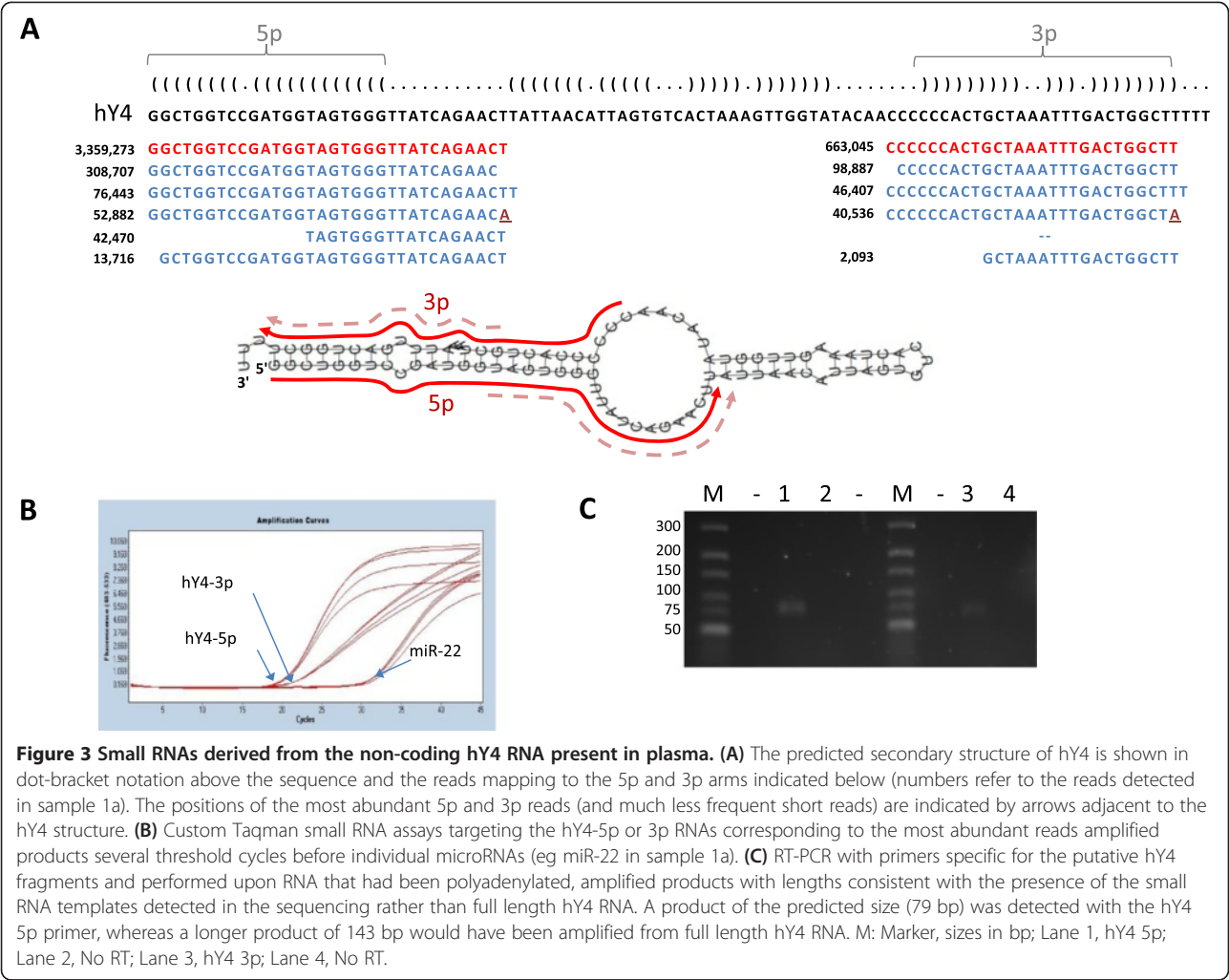
The contigs assigned to *Hypocreales* are extremely similar to the published sequences. For example, contig 44 has a similarity of 98.6% identity over 1162 nucleotides to *Hypocreales Cordycipitaceae Cordyceps gunnii* 28S ribosomal RNA (Figure 7A). This contig can also be aligned, with lower similarity, to rRNA from many other species. A region of contig 44 across which many orthologous sequences were available was selected and a multiple alignment made (Figure 7B). The phylogram derived from this alignment illustrates that contig 44 is considerably more similar to sequences from several species within *Hypocreales* than to those within *Malasseziales* and even more dissimilar to the human rRNA sequence (Figure 7C). Contigs generated from analysis of samples

from the study by Wang et al. [25] were also similar to fungal sequences and indeed some were identical to contig 44 for >700 bp (Additional file 5: Figure S3).

All the most abundant contigs fall within the mature rRNA regions but the distribution of detected reads is very uneven (Figure 8). Although the variation in coverage could be partially due to experimental bias (ie differential cloning efficiency of sequences [35]) it is also likely to reflect *in vivo* abundances.

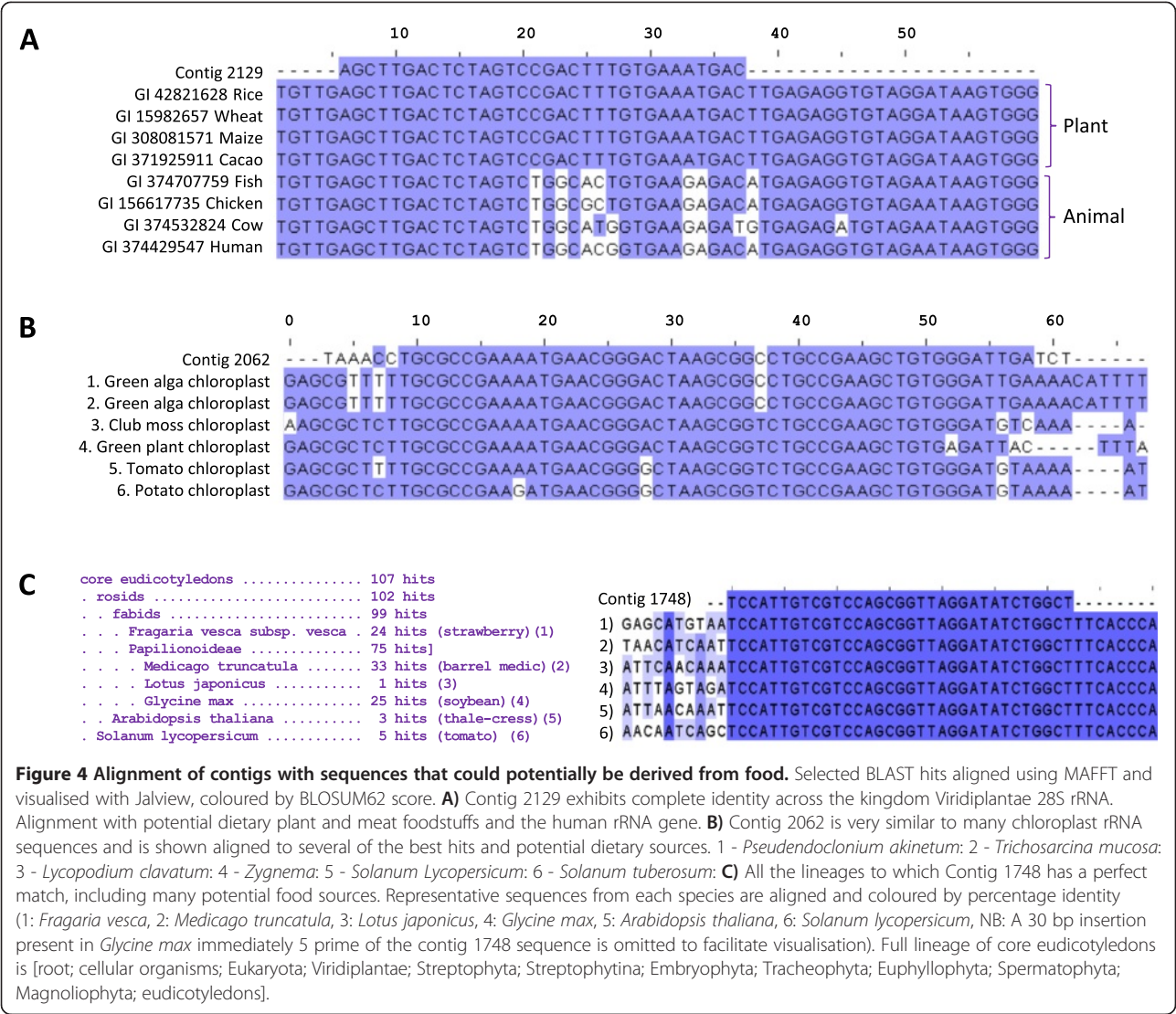
Discussion

Highly expressed small RNAs derived from Y RNAs hY1 and hY3 have been reported in tumours and high expression in serum suggested by RT-PCR [36]. We also observed a small number of sequences matching hY1 and hY3, but the presence of extremely abundant hY4



fragments, confirmed by RT-qPCR, was unexpected. Our ability to detect Y RNA fragments as such a large proportion of total small RNAs in this study may relate to practical details of the library preparation protocol employed, particularly the size range selected. Y RNAs form part of the RoRNP, which also contains the proteins Ro60 and La, but their function is poorly understood [37]. They are required for chromosomal replication [38] and are over-expressed in tumours [39]. It has been demonstrated that double-stranded RNA oligonucleotides comprising the stem of the Y RNA are sufficient to reconstitute DNA replication *in vitro* [40]. Y RNAs are rapidly degraded during apoptosis to generate fragments similar in size to those observed in this study [41]. Although it has been suggested that small RNAs derived from Y RNAs may act analogously to microRNAs, the formation of Y3 and Y5 RNA fragments has been shown to be Dicer independent [29]. Given the abundance of the hY4 fragments in plasma, it is an intriguing possibility that they may have some, as yet unknown function.

The detection of microbial sequences in plasma supports previous reports of circulating enterobacterial transcripts [27] and the most detailed study of these sequences to date by Wang *et al.* [25], who performed extensive control experiments to rule out potential sources of contamination. However, the possibility that observations of exogenous RNAs result from contamination remains a serious concern [42]. Spurious detection of such sequences could arise due to contamination during sample handling, library preparation or sequencing or result from errors in data analysis. It is difficult to envisage how contamination with identical sequences could occur in studies undertaken in diverse locations by independent investigators, ie as detected in this study and by Huang *et al.* [28] and Wang *et al.* [25] (Additional file 5: Figure S3). In addition, analysis of data from the sequence runs prior to those reported in this study confirmed that they were not the source of contamination. In this study reads were assembled to try to improve mapping accuracy and reduce the computational requirements for database

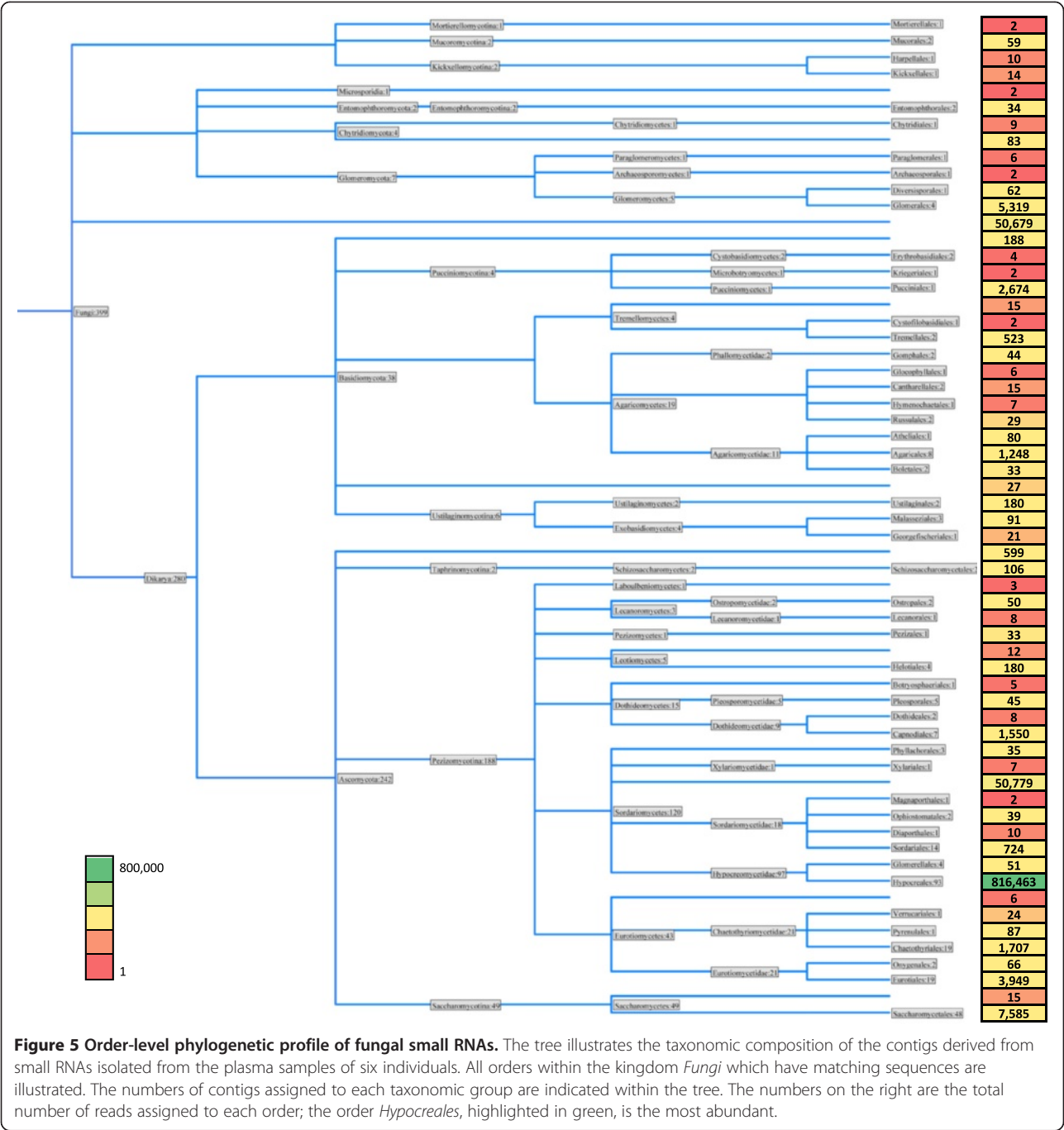


searching. The observation of similar mapping results without assembly of the sequence reads (Additional file 6: Figure S4) supports the proposed phylogenetic origins.

The taxonomic breakdown of the originating organisms achieved with our contig-based strategy is in broad agreement with that reported by Wang *et al.*; *Proteobacteria* were the most abundant bacterial phylum in both studies, with *Bacteroidetes* also commonly detected, whilst *Ascomycota* was the most abundant phylum of Fungi in both studies. However, our data suggest an even greater predominance of *Ascomycota* and we can assign many of these reads down to the level of Order (*Hypocreales*). Whilst members of this order have occasionally been reported as opportune pathogens in immunocompromised patients [43], they are more commonly plant or insect parasites [44], while *Hypocrea jecorina* is a widely used source of cellulases [45]. It is remarkable that the vast majority of fungal reads should be derived from a small number of

closely related species or potentially even a single species. From where do all these sequences originate?

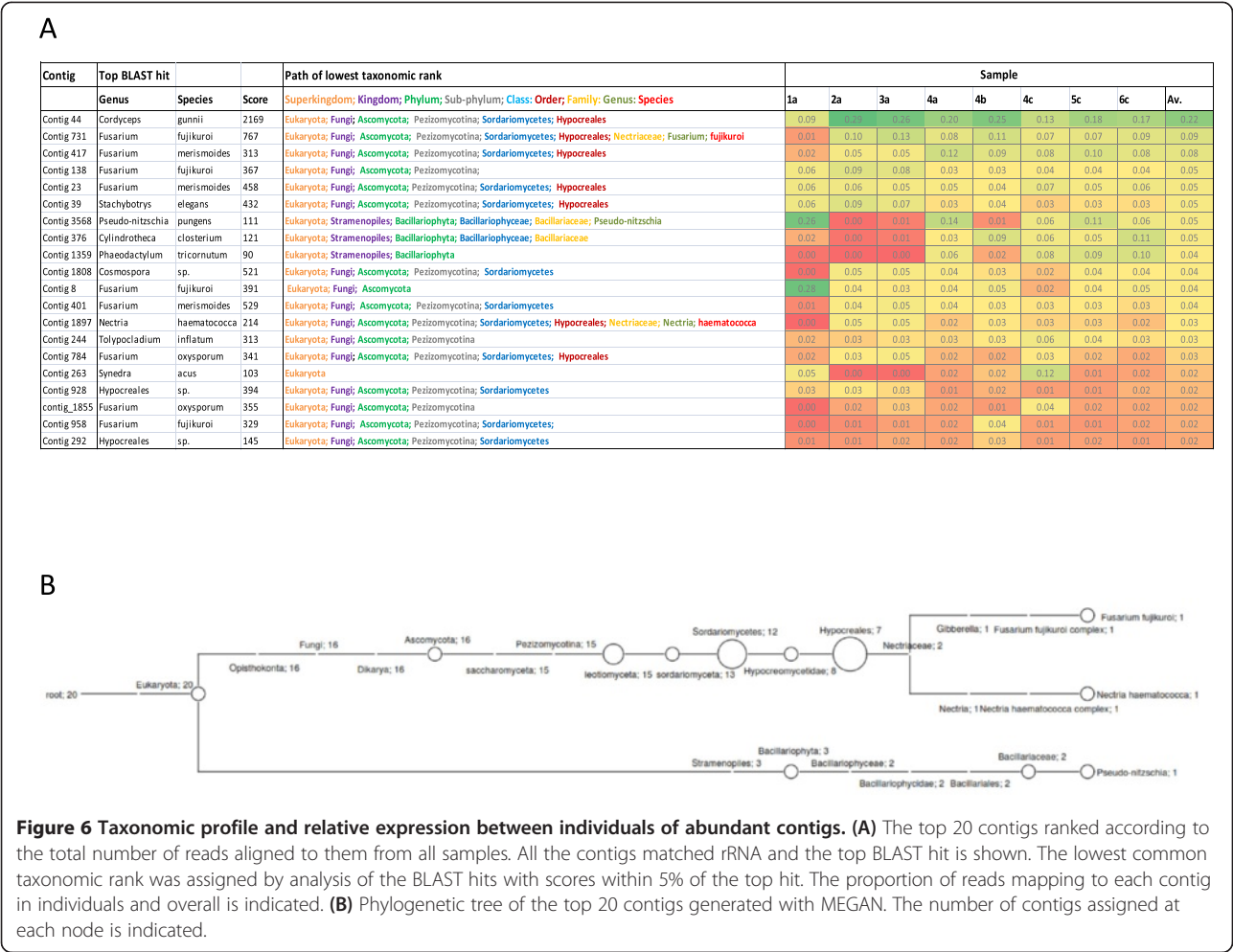
The composition of both the fungal and bacterial plasma microbiome detected suggests that the sequences do not result from contamination from the skin microbiome during collection of blood samples. Whilst the human skin microbiome varies widely, it is dominated by the bacterial phylum *Actinobacteria* (and to a lesser degree *Firmicutes* and *Proteobacteria*) [1,5] and the fungal genus *Malassezia* of the *Basidiomycota* phylum [6]. Reads from *Actinobacteria* comprised an average of 1.5% percent of bacterial reads in 5 of 6 samples and only 17.6% in the remaining sample. *Firmicutes* averaged 1% percent across all samples, although *Proteobacteria* were the most abundant (50%). With regard to fungi, only 3 contigs (91 reads) were assigned to *Malassezia*. It seems unlikely that contamination during sample processing could result in such similar microbiome profiles in three



independent plasma small RNA datasets and across multiple library preparation methods.

Small RNA sequences have been reported to enter the circulation from the gastrointestinal tract [31] and pharmacological preparations of small interfering RNAs (siRNAs) have been demonstrated to cross the gut wall following oral administration [46-48]. The gut therefore seems the most likely origin for microbial plasma small RNAs. The human gut, in contrast to skin, is predominantly colonised by the bacterial phyla *Bacteroidetes*

and *Firmicutes* [1,5], and by the fungal phylum *Ascomycota* [4]. It is therefore conceivable that the gut is the source, but one would not expect the observed predominance of sequences from *Hypocreales*. Perhaps the niche occupied by these species within the gut predisposes them to uptake into the circulation. The respiratory tract is another potential source and indeed *Fusarium* is one of the four most common pathogenic fungi detected, along with *Candida*, *Aspergillus* and *Cryptococcus* [49]. Although some microRNAs may be absorbed from the gut



unshielded to survive exposed in the circulation for several hours [31,50], many are protected from degradation by association with lipids and proteins [51,52] and there is some evidence that the exogenous RNAs may be similarly protected [25]. Indeed rRNA fragments have been shown to enter argonaute protein complexes [53]. Differential stability could contribute to over-representation of certain sequences.

In addition to RNAs of microbial origin, some sequences potentially derived from foodstuffs were detected. Notably the greatest proportion of reads matching plant sequences were found in sample 3, which was obtained from the one individual who reported following a vegetarian diet. Although it has been reported that plant microRNAs (xenomiRs) are not reliably detected in plasma after ingestion [54,55] the possibility of genetic material from food entering the circulation is supported by the detection of plant chloroplast DNA in the blood of cows [56]. The unequivocal assignment of significant numbers of circulating small RNAs to plant rRNA in this

study raises the exciting possibility that it may be possible to quantify diet from a simple blood test.

Great care must be taken when comparing between studies because differences in sample collection and library preparation can have profound effects upon the small RNA profiles observed and the proportion of reads mapping to Y RNAs or exogenous small RNAs. Nonetheless, the detection of these same small RNAs in diverse studies confirms that they are a common feature of the circulation.

Conclusion

Abundant fragments derived from the non-coding hY4 RNA, but of unknown function, have been detected in human plasma. RNAs from a diverse range of microbes are also present, but the majority of fungal sequences are from species in the Order *Hypocreales*. This raises questions about how these exogenous RNAs reach the circulation, whether they are functional and why specific fungi are so highly represented. This work has

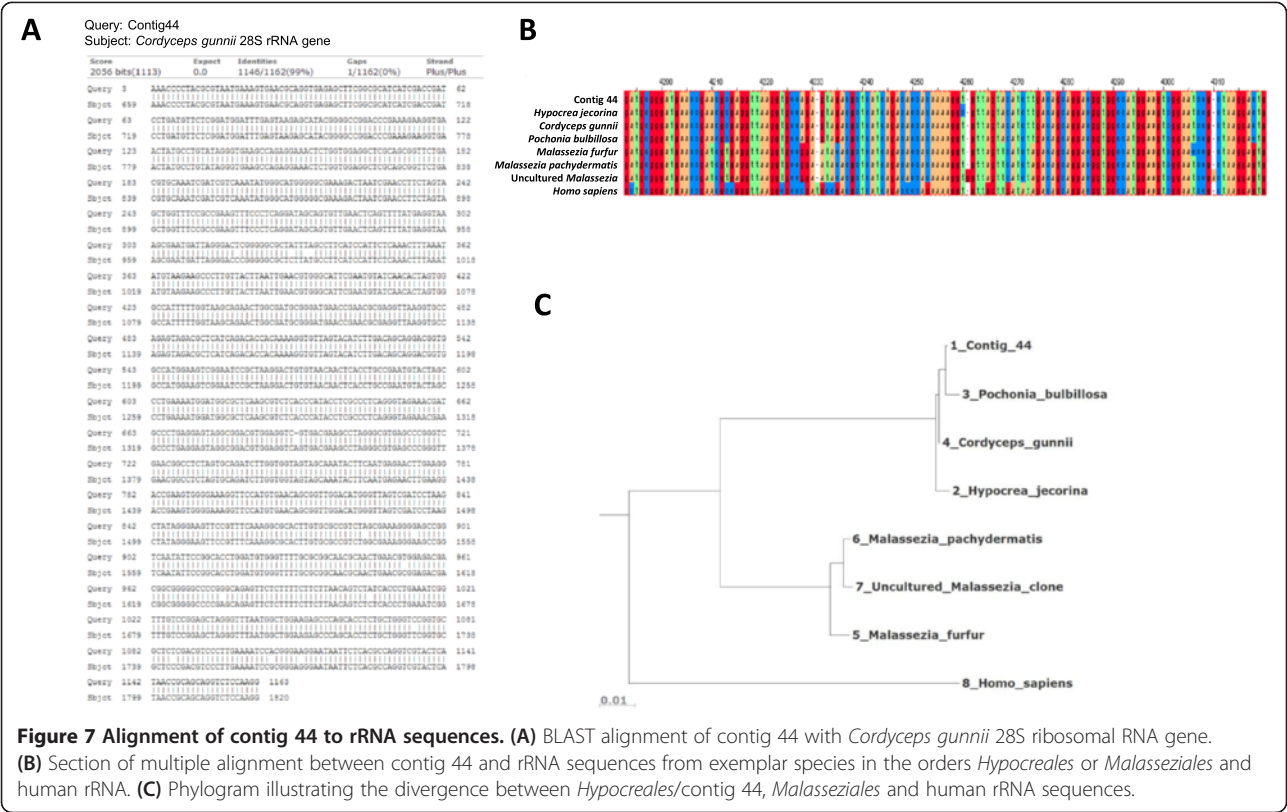


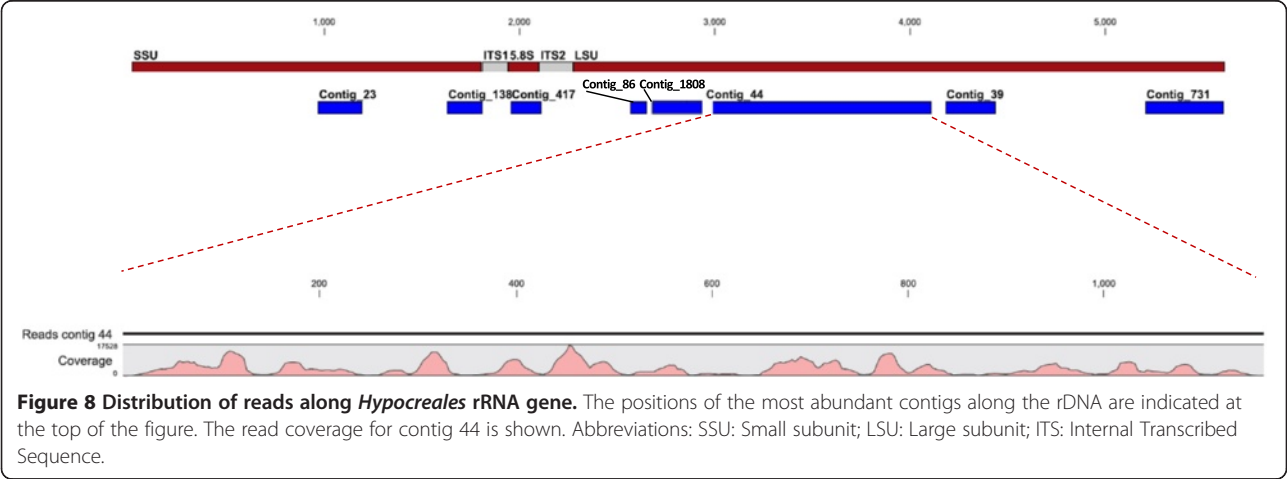
Figure 7 Alignment of contig 44 to rRNA sequences. (A) BLAST alignment of contig 44 with *Cordyceps gunnii* 28S ribosomal RNA gene. **(B)** Section of multiple alignment between contig 44 and rRNA sequences from exemplar species in the orders *Hypocreales* or *Malasseziales* and human rRNA. **(C)** Phylogram illustrating the divergence between *Hypocreales*/contig 44, *Malasseziales* and human rRNA sequences.

demonstrated the feasibility of determining the microbiome that contributes small RNAs to the blood. The profile of microbial sequences detected is almost certainly influenced by the composition of the wider microbiome, particularly in the gut. Given the integral role of the human microbiome in normal health and pathology, it seems likely that knowledge of the plasma microbiome will be soon prove to be of clinical importance.

Methods

Sample collection and RNA extraction

Three healthy individuals aged 20–40 years were recruited from Belfast, N. Ireland, UK: male, Caucasian (sample 1); female, Caucasian (sample 2); and male, Indian (sample 3). All participants completed a food-frequency questionnaire which included questions on any special dietary requirements. A blood sample was taken in EDTA-treated tubes



and plasma was separated immediately by centrifugation for 10 minutes at 1,000 g and subsequently at 10,000 g for 10 minutes prior to RNA extraction using a miRNeasy kit (Qiagen, Crawley, UK). RNA purity and quantity were determined using a Nanodrop spectrophotometer (Thermo Scientific) and Qubit fluorimeter (Life Technologies). RNA integrity was assessed using RNA 2000 and small RNA chips on a Bioanalyzer (Agilent).

Ethics and consent

This study was conducted according to the guidelines laid down in the Declaration of Helsinki and all procedures involving human participants/patients were approved by the Research Ethics Committee of the School of Medicine and Dentistry, Queen's University Belfast (Ref:11/05v3). Written informed consent was obtained from all participants.

Deep sequencing

Small RNA libraries were prepared using a Truseq small RNA sample prep kit (Illumina) following the manufacturer's protocol. This included size selection using a 6% PAGE Gel; the region between the custom Illumina markers was excised, corresponding to insert sizes of approximately 20–35 nucleotides. Cluster generation and sequencing with 40 nucleotide reads on a MiSeq was performed at the Trinity Genome Sequencing Laboratory, Dublin [57].

Data analysis

Sequencing data were analyzed using Genomics workbench software v5.5.1 (CLCbio, Aarhus, Denmark). After removal of adapter sequences, reads >15 bp and with at least 2 copies were aligned, allowing 2 mismatches, to miRBase (Release 19), a database of human non-coding RNA downloaded from Ensembl using Biomart [58] and the human genome (hg19). The remaining unannotated reads were pooled and assembled into contigs using the de novo assembly algorithm of Genomics workbench. Reads from each individual sample were then mapped back to the contigs. For subsequent phylogenetic analyses the putative origins of contig sequences were assigned using the sequence identifier (gi) numbers of the top hits determined by megablast [59,60] (available online [61]) against the NCBI non-redundant database (E-value <0.001). Lists of gi numbers were uploaded to the metagenomic analysis tools [62] available through the Galaxy platform [63,64], specifically to 'Fetch taxonomic representation', 'Summarize taxonomy', 'draw phylogeny' and 'Find lowest diagnostic rank'. Microsoft Access databases were used to integrate datasets. Taxonomic classification of the top 5% of BLAST hits was performed using the MEtaGenome ANalyzer (MEGAN) analysis tool [32,33]. The lowest common ancestor was assigned

following manual removal of individual hits with obviously incorrect taxonomic classifications (ie matching the query and top blast hits but not other sequences from their alleged species). Optimal RNA secondary structures were predicted using the Vienna RNAfold webserver [65,66]. Additional multiple sequence alignments were performed using the Multiple Alignment using Fast Fourier Transform (MAFFT) program [67] available online [68] or Clustal Omega [69,70], available through the EBI server [71]. Multiple alignments were visualised with Jalview [72] and phylograms with Archaeopteryx [73]. Custom Perl scripts were used for manipulating sequence files.

RT-PCR

Y-RNA custom small RNA Taqman assays (Life Technologies) were designed to target the following sequences: HY4_5p; GGCUGGUCCGAUGGUAGUGGGUUAUCAG AACU and HY4_3p; CCCCCACUGCUAAAUUUGA CUGGCUU. Taqman reverse transcription and PCR were performed according to the manufacturer's instructions on a LightCycler480 platform (Roche).

For detection of Y-RNA fragments, RNA was polyadenylated using *E. coli* Poly(A) Polymerase I (Ambion) and reverse transcribed using Super Script III reverse transcriptase (Life Technologies) and an oligo-dt adaptor: GCGAGCACAGAATTAATACGACTCACTATAGGTTT TTTTTTTTTVN. PCR was performed using the common reverse primer GCGAGCACAGAATTAATACGAC and either an HY4_5p primer: GGCTGGTCCGATGG TAGT or HY4_3p primer: CCCCCACTGCTAAAA TTTGA. 35 cycles of PCR were performed with the following conditions 94°C 30 sec; 56°C 30 sec; 72°C 1 minute using Hotstar Taq DNA polymerase (Qiagen).

Availability of supporting data

The data sets supporting the results of this article are available in the Gene Expression Omnibus (GEO) repository [74]. The sequencing data generated in this study has accession number GSE52981 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE52981>) and the publicly available plasma small RNA sequencing data [28] analysed has accession number GSE45722.

Additional files

Additional file 1: Excel spreadsheet. Number of reads from each sample annotated by gene type or phylogenetic category.

Additional file 2: Figure S1. Distribution of read lengths in a range of sequencing libraries prepared from blood. The percentages of reads of each length are shown for libraries prepared from plasma, exosomes isolated from plasma or whole blood, including cells. Both the source material and library preparation protocol (eg size selection) influence the insert sizes observed. References: Huang et al. [28]; Wang et al. [25]; Leidinger et al. [30].

Additional file 3: Excel spreadsheet. Full lineages of the top blast hits of all contigs.

Additional file 4: Figure S2. Multiple alignment of Contig_1808 with fungal rRNAs. Contig 1808 is aligned with the five most similar sequences in the NCBI nr database and the rRNA sequence of *Fusarium venenatum*.

Additional file 5: Figure S3. Analysis of plasma sequence data from Wang et al. [25]. (a) Taxonomic composition of the contigs derived from small RNAs isolated from a normal plasma sample (ERR248695), determined from BLAST searches using MEGAN. (b) Alignment of one contig derived from sample ERR248695 from the study by Wang et al. [25] with contig 44 from this study, demonstrating total identity.

Additional file 6: Figure S4. Phylogenetic profiles predicted from individual reads or contigs. A random subset of the reads that were unannotated to human databases was generated from Sample 3A. These were input either directly or after assembly into contigs to BLAST searches of the nt database. Similarities with fungal sequences are a key feature detected by both approaches. Using this subset of sequences no contigs with potential bacterial origin were detected, probably reflecting the relatively low abundance of putative bacterial reads in this sample in comparison to fungal reads (see Figure 2D). (a) Phylogenetic profile predicted using MEGAN to interpret BLAST searches using contigs assembled from the reads. The number of hits at each node is indicated. (b) Phylogenetic profile predicted from BLAST searches of individual reads. The similarity between the trees suggests that mapping of assembled reads is broadly consistent to the results with individual reads.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

MB carried out the bioinformatic analyses of sequence data, participated in the design of the study and helped to draft the manuscript. JG prepared the libraries and helped to draft the manuscript. EB performed RT-PCR analyses. SB wrote perl scripts and provided comments. UC participated in the design of the study. RH helped to design the study and draft the manuscript. DS conceived and designed the study, led and performed data analysis and drafted the manuscript. All authors read and approved the final manuscript.

Acknowledgements

Thanks to Estelle Lowry for blood collection and Jayne Woodside and Margaret Dellett for helpful comments and discussion. This work was funded by the Biotechnology and Biological Sciences Research Council (BBSRC) (Grant number: BB/H005498/1) and the Department for Employment and Learning, Northern Ireland.

Received: 25 April 2014 Accepted: 16 October 2014

Published: 25 October 2014

References

- Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, Nielsen T, Pons N, Levenez F, Yamada T, Mende DR, Li J, Xu J, Li S, Li D, Cao J, Wang B, Liang H, Zheng H, Xie Y, Tap J, Lepage P, Bertalan M, Batto JM, Hansen T, Le Paslier D, Linneberg A, Nielsen HB, Pelletier E, Renault P, et al: **A human gut microbial gene catalogue established by metagenomic sequencing.** *Nature* 2010, **464**(7285):59–65.
- Wang ZK, Yang YS: **Upper gastrointestinal microbiota and digestive diseases.** *World J Gastroenterol* 2013, **19**(10):1541–1550.
- Human Microbiome Project (HMP). <http://commonfund.nih.gov/Hmp/>.
- Ott SJ, Kuhbacher T, Musfeldt M, Rosenstiel P, Hellmig S, Rehman A, Drews O, Weichert W, Timmis KN, Schreiber S: **Fungi and inflammatory bowel diseases: Alterations of composition and diversity.** *Scand J Gastroenterol* 2008, **43**(7):831–841.
- Cho I, Blaser MJ: **The human microbiome: at the interface of health and disease.** *Nat Rev Genet* 2012, **13**(4):260–270.
- Findley K, Oh J, Yang J, Conlan S, Deming C, Meyer JA, Schoenfeld D, Nomicos E, Park M, NIH Intramural Sequencing Center Comparative Sequencing Program, Kong HH, Segre JA: **Topographic diversity of fungal and bacterial communities in human skin.** *Nature* 2013, **498**(7454):367–370.
- Faith JJ, Guruge JL, Charbonneau M, Subramanian S, Seedorf H, Goodman AL, Clemente JC, Knight R, Heath AC, Leibel RL, Rosenbaum M, Gordon JL: **The long-term stability of the human gut microbiota.** *Science* 2013, **341**(6141):1237439.
- Honda K, Littman DR: **The microbiome in infectious disease and inflammation.** *Annu Rev Immunol* 2012, **30**:759–795.
- Hoffmann C, Dollive S, Grunberg S, Chen J, Li H, Wu GD, Lewis JD, Bushman FD: **Archaea and fungi of the human gut microbiome: correlations with diet and bacterial residents.** *PLoS One* 2013, **8**(6):e66019.
- Simren M, Barbara G, Flint HJ, Spiegel BM, Spiller RC, Vanner S, Verdu EF, Whorwell PJ, Zoetendal EG, Rome Foundation Committee: **Intestinal microbiota in functional bowel disorders: a Rome foundation report.** *Gut* 2013, **62**(1):159–176.
- Rigsbee L, Agans R, Shankar V, Kenche H, Khamis HJ, Michail S, Paliy O: **Quantitative profiling of gut microbiota of children with diarrhea-predominant irritable bowel syndrome.** *Am J Gastroenterol* 2012, **107**(11):1740–1751.
- Ley RE, Turnbaugh PJ, Klein S, Gordon JL: **Microbial ecology: human gut microbes associated with obesity.** *Nature* 2006, **444**(7122):1022–1023.
- Greenblum S, Turnbaugh PJ, Borenstein E: **Metagenomic systems biology of the human gut microbiome reveals topological shifts associated with obesity and inflammatory bowel disease.** *Proc Natl Acad Sci U S A* 2012, **109**(2):594–599.
- Zhao L: **The gut microbiota and obesity: from correlation to causality.** *Nat Rev Microbiol* 2013, **11**:639–647.
- Koeth RA, Wang Z, Levison BS, Buffa JA, Org E, Sheehy BT, Britt EB, Fu X, Wu Y, Li L, Smith JD, DiDonato JA, Chen J, Li H, Wu GD, Lewis JD, Warrier M, Brown JM, Krauss RM, Tang WH, Bushman FD, Lusis AJ, Hazen SL: **Intestinal microbiota metabolism of L-carnitine, a nutrient in red meat, promotes atherosclerosis.** *Nat Med* 2013, **19**(5):576–585.
- Koren O, Spor A, Felin J, Fak F, Stombaugh J, Tremaroli V, Behre CJ, Knight R, Fagerberg B, Ley RE, Backhed F: **Human oral, gut, and plaque microbiota in patients with atherosclerosis.** *Proc Natl Acad Sci U S A* 2011, **108**(Suppl 1):4592–4598.
- Backhed F: **Meat-metabolizing bacteria in atherosclerosis.** *Nat Med* 2013, **19**(5):533–534.
- Karlsson FH, Tremaroli V, Nookaew I, Bergstrom G, Behre CJ, Fagerberg B, Nielsen J, Backhed F: **Gut metagenome in European women with normal, impaired and diabetic glucose control.** *Nature* 2013, **498**(7452):99–103.
- Wu HJ, Ivanov II, Darce J, Hattori K, Shima T, Umesaki Y, Littman DR, Benoist C, Mathis D: **Gut-residing segmented filamentous bacteria drive autoimmune arthritis via T helper 17 cells.** *Immunity* 2010, **32**(6):815–827.
- Scher JU, Szczesnak A, Longman RS, Segata N, Ubeda C, Bielski C, Rostron T, Cerundolo V, Pamer EG, Abramson SB, Huttenhower C, Littman DR: **Expansion of intestinal Prevotella copri correlates with enhanced susceptibility to arthritis.** *Elife* 2013, **2**(0). doi:10.7554/eLife.01202.
- Hsiao EY, McBride SW, Hsien S, Sharon G, Hyde ER, McCue T, Codelli JA, Chow J, Reisman SE, Petrosino JF, Patterson PH, Mazmanian SK: **Microbiota modulate behavioral and physiological abnormalities associated with neurodevelopmental disorders.** *Cell* 2013, **155**(7):1451–1463.
- Iliev ID, Funari VA, Taylor KD, Nguyen Q, Reyes CN, Strom SP, Brown J, Becker CA, Fleshner PR, Dubinsky M, Rotter JI, Wang HL, McGovern DP, Brown GD, Underhill DM: **Interactions between commensal fungi and the C-type lectin receptor Dectin-1 influence colitis.** *Science* 2012, **336**(6086):1314–1317.
- Geekianage H, Jicha GA, Nelson PT, Chan C: **Blood serum miRNA: non-invasive biomarkers for Alzheimer's disease.** *Exp Neurol* 2012, **235**(2):491–496.
- Guay C, Roggli E, Nesca V, Jacovetti C, Regazzi R: **Diabetes mellitus, a microRNA-related disease?** *Transl Res* 2011, **157**(4):253–264.
- Wang K, Li H, Yuan Y, Etheridge A, Zhou Y, Huang D, Wilmes P, Galas D: **The complex exogenous RNA spectra in human plasma: an interface with human gut biota?** *PLoS One* 2012, **7**(12):e51009.
- Semenov DV, Baryakin DN, Kamynina TP, Kuligina EV, Richter VA: **Fragments of noncoding RNA in plasma of human blood.** *Ann N Y Acad Sci* 2008, **1137**:130–134.
- Semenov DV, Baryakin DN, Brenner EV, Kurilshikov AM, Vasiliev GV, Bryzgalov LA, Chikova ED, Filippova JA, Kuligina EV, Richter VA: **Unbiased approach to profile the variety of small non-coding RNA of human blood plasma with massively parallel sequencing technology.** *Expert Opin Biol Ther* 2012, **12**(Suppl 1):S43–S51.

28. Huang X, Yuan T, Tschannen M, Sun Z, Jacob H, Du M, Liang M, Dittmar RL, Liu Y, Liang M, Kohli M, Thibodeau SN, Boardman L, Wang L: **Characterization of human plasma-derived exosomal RNAs by deep sequencing.** *BMC Genomics* 2013, **14**:319.
29. Nicolas FE, Hall AE, Csorba T, Turnbull C, Dalmay T: **Biogenesis of Y RNA-derived small RNAs is independent of the microRNA pathway.** *FEBS Lett* 2012, **586**(8):1226–1230.
30. Leidinger P, Backes C, Deutscher S, Schmitt K, Muller SC, Frese K, Haas J, Ruprecht K, Paul F, Stahler C, Lang CJ, Meder B, Bartfai T, Meese E, Keller A: **A blood based 12-miRNA signature of Alzheimer disease patients.** *Genome Biol* 2013, **14**(7):R78.
31. Zhang L, Hou D, Chen X, Li D, Zhu L, Zhang Y, Li J, Bian Z, Liang X, Cai X, Yin Y, Wang C, Zhang T, Zhu D, Zhang D, Xu J, Chen Q, Ba Y, Liu J, Wang Q, Chen J, Wang J, Wang M, Zhang Q, Zhang J, Zen K, Zhang CY: **Exogenous plant MIR168a specifically targets mammalian LDLRAP1: evidence of cross-kingdom regulation by microRNA.** *Cell Res* 2012, **22**(1):107–126.
32. Huson DH, Auch AF, Qi J, Schuster SC: **MEGAN analysis of metagenomic data.** *Genome Res* 2007, **17**(3):377–386.
33. Huson DH, Mitra S, Ruscheweyh HJ, Weber N, Schuster SC: **Integrative analysis of environmental sequences using MEGAN4.** *Genome Res* 2011, **21**(9):1552–1560.
34. O'Donnell K, Cigelnik E, Casper HH: **Molecular phylogenetic, morphological, and mycotoxin data support reidentification of the Quorn mycoprotein fungus as Fusarium venenatum.** *Fungal Genet Biol* 1998, **23**(1):57–67.
35. Sorefan K, Pais H, Hall AE, Kozomara A, Griffiths-Jones S, Moulton V, Dalmay T: **Reducing ligation bias of small RNAs in libraries for next generation sequencing.** *Silence* 2012, **3**(1):4. -907X-3-4.
36. Meiri E, Levy A, Benjamin H, Ben-David M, Cohen L, Dov A, Dromi N, Elyakim E, Yerushalmi N, Zion O, Litwick-Yanai G, Sitbon E: **Discovery of microRNAs and other small RNAs in solid tumors.** *Nucleic Acids Res* 2010, **38**(18):6234–6246.
37. Kohn M, Pazaitis N, Huttelmaier S: **Why YRNAs? About versatile RNAs and their functions.** *Biogeosciences* 2013, **3**(1):143–156.
38. Christov CP, Gardiner TJ, Szuts D, Krude T: **Functional requirement of noncoding Y RNAs for human chromosomal DNA replication.** *Mol Cell Biol* 2006, **26**(18):6993–7004.
39. Christov CP, Trivier E, Krude T: **Noncoding human Y RNAs are overexpressed in tumours and required for cell proliferation.** *Br J Cancer* 2008, **98**(5):981–988.
40. Gardiner TJ, Christov CP, Langley AR, Krude T: **A conserved motif of vertebrate Y RNAs essential for chromosomal DNA replication.** *RNA* 2009, **15**(7):1375–1385.
41. Rutjes SA, van der Heijden A, Utz PJ, van Venrooij WJ, Pruijn GJ: **Rapid nucleolytic degradation of the small cytoplasmic Y RNAs during apoptosis.** *J Biol Chem* 1999, **274**(35):24799–24807.
42. Zhang Y, Wiggins BE, Lawrence C, Petrick J, Ivashuta S, Heck G: **Analysis of plant-derived miRNAs in animal small RNA datasets.** *BMC Genomics* 2012, **13**:381.
43. Howard DH: *Pathogenic Fungi in Humans and Animals (Mycology).* 2nd edition. Boca Raton, FL, USA: CRC Press; 2002.
44. Berbee ML: **The phylogeny of plant and animal pathogens in the Ascomycota.** *Physiol Mol Plant Pathol* 2001, **59**(4):165–187.
45. Martinez D, Berka RM, Henrissat B, Saloheimo M, Arvas M, Baker SE, Chapman J, Chertkov O, Coutinho PM, Cullen D, Danchin EG, Grigoriev IV, Harris P, Jackson M, Kubicek CP, Han CS, Ho I, Larrondo LF, de Leon AL, Magnuson JK, Merino S, Misra M, Nelson B, Putnam N, Robertse B, Salamov AA, Schmolli M, Terry A, Thayer N, Westerholm-Parvinen A, et al: **Genome sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn. *Hypocrea jecorina*).** *Nat Biotechnol* 2008, **26**(5):553–560.
46. Aouadi M, Tesz GJ, Nicoloso SM, Wang M, Chouinard M, Soto E, Ostroff GR, Czech MP: **Orally delivered siRNA targeting macrophage Map4k4 suppresses systemic inflammation.** *Nature* 2009, **458**(7242):1180–1184.
47. Xu J, Ganesh S, Amiji M: **Non-condensing polymeric nanoparticles for targeted gene and siRNA delivery.** *Int J Pharm* 2012, **427**(1):21–34.
48. Akhtar S: **Oral delivery of siRNA and antisense oligonucleotides.** *J Drug Target* 2009, **17**(7):491–495.
49. Ghannoum MA, Jurevic RJ, Mukherjee PK, Cui F, Sikaroodi M, Naqvi A, Gillevet PM: **Characterization of the oral fungal microbiome (mycobiome) in healthy individuals.** *PLoS Pathog* 2010, **6**(1):e1000713.
50. Witwer KW: **XenomiRs and miRNA homeostasis in health and disease: evidence that diet and dietary miRNAs directly and indirectly influence circulating miRNA profiles.** *RNA Biol* 2012, **9**(9):1147–1154.
51. Arroyo JD, Chevillet JR, Kroh EM, Ruf IK, Pritchard CC, Gibson DF, Mitchell PS, Bennett CF, Pogonosov-Agadjanyan EL, Stirewalt DL, Tait JF, Tewari M: **Argonaute2 complexes carry a population of circulating microRNAs independent of vesicles in human plasma.** *Proc Natl Acad Sci U S A* 2011, **108**(12):5003–5008.
52. Vickers KC, Palmisano BT, Shoucri BM, Shamburek RD, Remaley AT: **MicroRNAs are transported in plasma and delivered to recipient cells by high-density lipoproteins.** *Nat Cell Biol* 2011, **13**(4):423–433.
53. Wei H, Zhou B, Zhang F, Tu Y, Hu Y, Zhang B, Zhai Q: **Profiling and identification of small rDNA-derived RNAs and their potential biological functions.** *PLoS One* 2013, **8**(2):e56842.
54. Witwer KW, McAlexander MA, Queen SE, Adams RJ: **Real-time quantitative PCR and droplet digital PCR for plant miRNAs in mammalian blood provide little evidence for general uptake of dietary miRNAs: limited evidence for general uptake of dietary plant xenomiRs.** *RNA Biol* 2013, **10**(7):1080–1086.
55. Snow JW, Hale AE, Isaacs SK, Baggish AL, Chan SY: **Ineffective delivery of diet-derived microRNAs to recipient animal organisms.** *RNA Biol* 2013, **10**(7):1107–1116.
56. Bertheau Y, Helbling JC, Fortabat MN, Makhzami S, Sotinel I, Audeon C, Nignol AC, Kobilinsky A, Petit L, Fach P, Brunschwig P, Duhem K, Martin P: **Persistence of plant DNA sequences in the blood of dairy cows fed with genetically modified (Bt176) and conventional corn silage.** *J Agric Food Chem* 2009, **57**(2):509–516.
57. Trinity Genome Sequencing Laboratory. <http://www.medicine.tcd.ie/sequencing>.
58. Biomart for the Ensembl database. <http://www.ensembl.org/biomart/martview/>.
59. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215**(3):403–410.
60. Morgulis A, Coulouris G, Raytselis Y, Madden TL, Agarwala R, Schaffer AA: **Database indexing for production MegaBLAST searches.** *Bioinformatics* 2008, **24**(16):1757–1764.
61. **Basic Local Alignment Search Tool (BLAST) home page at NCBI.** <http://blast.ncbi.nlm.nih.gov/>.
62. Kosakovsky Pond S, Wadhwani S, Chiaromonte F, Ananda G, Chung WY, Taylor J, Nekrutenko A, Galaxy Team: **Windshield splatter analysis with the Galaxy metagenomic pipeline.** *Genome Res* 2009, **19**(11):2144–2153.
63. **The Galaxy project.** <http://usegalaxy.org>.
64. Goecks J, Nekrutenko A, Taylor J, Galaxy Team: **Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences.** *Genome Biol* 2010, **11**(8):R86.
65. Gruber AR, Lorenz R, Bernhart SH, Neubock R, Hofacker IL: **The Vienna RNA websuite.** *Nucleic Acids Res* 2008, **36**(Web Server issue):W70–W74.
66. **Vienna RNAfold webserver.** <http://rna.tbi.univie.ac.at>.
67. Katoh K, Misawa K, Kuma K, Miyata T: **MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform.** *Nucleic Acids Res* 2002, **30**(14):3059–3066.
68. **Multiple Alignment using Fast Fourier Transform (MAFFT).** <http://mafft.cbrc.jp/alignment/server/>.
69. Goujon M, McWilliam H, Li W, Valentin F, Squizzato S, Paern J, Lopez R: **A new bioinformatics analysis tools framework at EMBL-EBI.** *Nucleic Acids Res* 2010, **38**(Web Server issue):W695–W699.
70. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Soding J, Thompson JD, Higgins DG: **Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega.** *Mol Syst Biol* 2011, **7**:539.
71. **Clustal Omega on the EBI server.** <http://www.ebi.ac.uk/Tools/msa/clustalo/>.
72. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ: **Jalview Version 2—a multiple sequence alignment editor and analysis workbench.** *Bioinformatics* 2009, **25**(9):1189–1191.
73. Zmasek CM, Eddy SR: **ATV: display and manipulation of annotated phylogenetic trees.** *Bioinformatics* 2001, **17**(4):383–384.
74. Barrett T, Troup DB, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Muerter RN, Holko M, Ayanbule O, Yefanov A, Soboleva A: **NCBI GEO: archive for functional genomics data sets—10 years on.** *Nucleic Acids Res* 2010, **39**(Database issue):D1005–D1010.

doi:10.1186/1471-2164-15-933

Cite this article as: Beatty et al.: Small RNAs from plants, bacteria and fungi within the order *Hypocreales* are ubiquitous in human plasma. *BMC Genomics* 2014 **15**:933.